IDC PERSPECTIVE

Maturing Hyperconverged Infrastructure Is Running Mission-Critical Workloads in Many Enterprises Today

Eric Burgener Carol Sliwa

EXECUTIVE SNAPSHOT

FIGURE 1

Executive Snapshot: HCI Becoming an Increasingly Viable Alternative for Mission-Critical Workloads

Hyperconverged infrastructure (HCI) is maturing rapidly and is being used to run workloads enterprises consider to be mission critical in more than half of the enterprises. As enterprises increasingly move in the direction of software-defined infrastructure, HCI is expected to increase its penetration in the enterprise. HCI is one of the fastest-growing enterprise storage market segments and is expected to grow from an $11.5 billion market in 2022 to a $14.0 billion market by the end of 2024.

Key Takeaways

- The key concerns that have kept HCI from being deployed with mission-critical workloads — an inability to scale compute and storage independently, node-level performance and capacity, and overall availability — are increasingly being addressed with new capabilities in the latest HCI offerings.
- 74.0% of enterprises have moved one or more SAN or NAS workloads from external arrays to HCI when those arrays have come up for technology refresh.
- With increasing use of HCI for mission-critical workloads, security has risen as a new concern for 44.6% of enterprises that are considering broader use of HCI. HCI vendors have responded with an array of encryption capabilities but there is still work to be done in this area that go beyond just encryption.

Recommended Actions

- When SAN and NAS arrays come up for technology refresh, consider software-defined storage options like HCI. When HCI can meet performance, availability, and other requirements, it can offer increased agility, much easier management and scaling, and better economics.
- Don't dismiss HCI as an option without understanding the updated capabilities of these platforms — they have made significant progress in their ability to support mission-critical enterprise workloads in just the past two years.
- Keep an eye on new developments in the "software-defined space," since new approaches such as composable/disaggregated infrastructure (CDI) are emerging that may ultimately provide very viable alternatives to popular software-defined storage platforms like HCI.

Source: IDC, 2022

June 2022, IDC #US49103022

SITUATION OVERVIEW

Facebook was looking for a better way to build web infrastructure in the 2005 time frame when it introduced the concept of hyperconverged infrastructure (HCI). The idea of storage software running on commodity off-the-shelf x86 server-based storage promised greater flexibility, improved ease of use, and better economics relative to the more traditional three-tier architectures. Hyperscalers in general followed Facebook's lead with HCI, and in the late 2000s, start-ups such as Nutanix, SimpliVity, Scale Computing, and Springpath sought to apply the unique architecture more generally in the enterprise market. Since then, there have been several acquisitions. EMC bought ScaleIO in 2013, and following Dell's acquisition of EMC, the HCI technology became PowerFlex. In 2017, Cisco purchased Springpath, and HPE acquired SimpliVity. Quantum added Pivot3 in 2021. VMware entered the market with VMware vSAN in 2014, and its HCI technology is widely adopted among small and medium-sized enterprises. IDC projects that HCI will be an $11.5 billion market in 2022 and grow to top $14.0 billion by 2024, when the overall external storage market will be $35.5 billion.

HCI software has matured significantly over the past decade, and recent survey data from IDC indicates that 74.0% of enterprises have moved one or more SAN or NAS workloads from external arrays to HCI. Enterprises often consider moving to HCI when legacy external arrays come up for technology refresh and 56.6% run mission-critical workloads on HCI. Many organizations use HCI for mixed workload consolidation of both mission-critical and non-mission-critical applications.

One of the traditional concerns with HCI technology has been the inability to scale compute and storage resources independently. Administrators had to scale HCI configurations by adding nodes with both compute and storage resources, forcing them to pay for unwanted storage capacity when they needed only compute and vice versa. That was one of the primary reasons that HCI clusters generally did not grow to a large scale. A 2021 IDC survey showed that 56% of enterprises have clusters with fewer than 10 nodes, and an additional 29% have 11–32 nodes, with the metric heavily weighted toward clusters of fewer than 16 nodes. For small and medium-sized enterprises, the advantages of HCI outweigh the inefficiency of excess resource allocation, which is not that costly for small configurations. Once workloads grow beyond 16 HCI nodes, customers often prefer a disaggregated architecture, which allows them to add and pay for only the compute, storage, and/or networking resources they specifically need.

HCI vendors recognized the need to scale compute and storage separately and introduced features to provide customers with more flexible resource allocation options. Most vendors offer compute-intensive and storage-intensive nodes that favor a particular resource type while still retaining the ease of scaling environments through simple node additions. Other vendors have introduced the ability to add cloud-based resources of a particular type that come close to the precision of resource allocation achieved by a disaggregated architecture. There may be latency implications for storage access in those configurations, so enterprises need to take that into account if they intend to make use of those capabilities. However, it's important for enterprises to realize that the historical concern about potentially inefficient resource allocation in large HCI configurations may be far less of an issue today, and HCI vendors support maximum HCI cluster sizes of 48 nodes and beyond.

Interestingly, HPE took a different approach to address this issue with the introduction of HPE Nimble Storage disaggregated HCI (dHCI) in 2020. The dHCI product combines HPE ProLiant servers with an HPE Nimble storage array factory configured and pre cabled into a rack under a single SKU. The product includes an HCI-like interface to allow administrators to manage the configuration like a traditional HCI platform but provides the option to independently scale compute by adding servers, or

storage, by adding Nimble arrays. In many ways, dHCI bears similarity to converged infrastructure offerings that have been available for years, but HPE markets it to appeal to an audience considering HCI, unlike the vendors of traditional converged infrastructure. HPE sells dHCI for large enterprise workloads where independent resource scaling has value, touting its ease of deployment, management, and configurability, while still selling SimpliVity, its more traditional HCI platform, for small and medium-sized enterprise workloads.

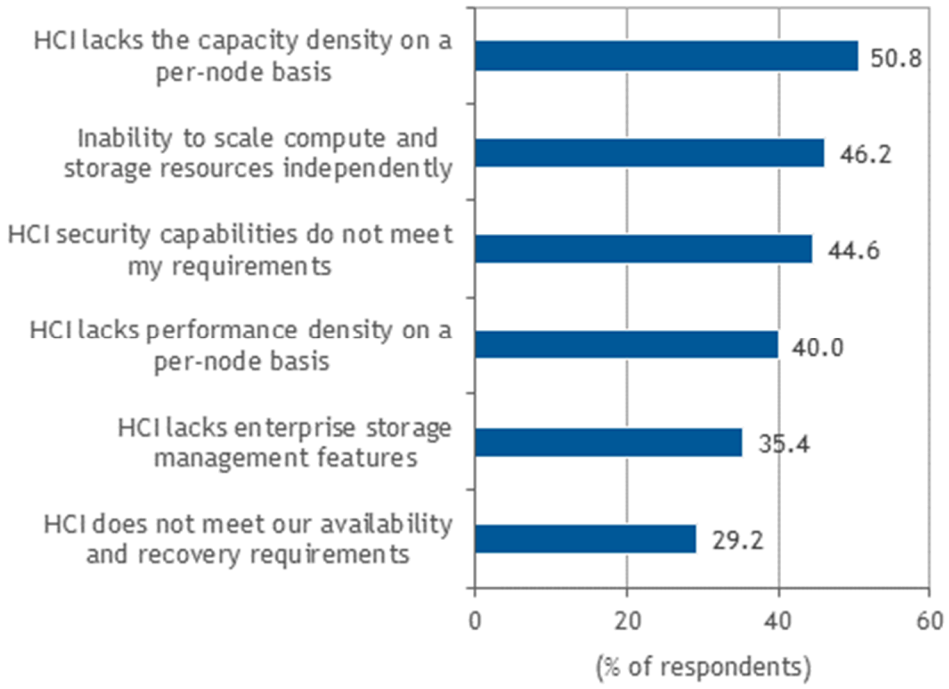
Mission-Critical Workloads and HCI

Enterprises have cited other concerns with HCI over the past decade with respect to certain workloads, including performance, availability (especially the recovery times associated with device and/or node failures), and single-node capacity. IDC's *HCI Survey,* completed in January 2022, shows those concerns still exist, although the order of importance has fluctuated and new ones, such as security, have surfaced. Survey respondents indicated which issues would prompt them to stay with their legacy SAN/NAS arrays for mission-critical workloads on technology refresh instead of looking at HCI (see Figure 2).

FIGURE 2

Reasons Not to Replace SAN/NAS with HCI

Q. If you have not replaced SAN/NAS with HCI, why not?



n = 258

Source: IDC's *HCI Survey,* January 2022

With the broader use of high-density solid state storage drives (SSDs), per-node capacity has been increasing in HCI environments. Some HCI vendors already have data placement algorithms that

strive to keep application data in the same HCI node that runs the application. These dynamic algorithms adjust as workloads evolve over time, continuously seeking to optimize data locality based on access patterns. With improving per-node capacity density, this concern about capacity density on a per-node basis should recede over time. According to IDC's 4Q21 *IT Infrastructure Survey,* released in February 2022, the weighted average for expected data growth in 2022 across all regions and company sizes is 55%.

Most HCI configurations use Ethernet as a cluster interconnect, and the latency associated with the network hop over TCP to access data outside of a particular node is another reason vendors implemented data locality algorithms. Faster interconnects such as Fibre Channel were too expensive to use in HCI environments. However, the maturation of the NVMe protocol and new NVMe over Fabrics options may help cost-effectively address the latency of internode access. NVMe over TCP (NVMe/TCP) offers latencies that are significantly lower than the traditional approaches used in HCI configurations, and its use incurs little if any additional cost. NVMe/TCP will run over standard converged ethernet adapters configured into most x86-based servers, and NVMe/TCP drivers are available for both Linux and Windows distributions. In fact, the NVMe/TCP driver is embedded into major commercial Linux distributions such as Red Hat Enterprise Linux. Once more HCI vendors start to support NVMe/TCP, it could help address the "capacity density on a per-node basis" concerns as well.

Security is a major concern for 44.6% of enterprises considering HCI. While security requirements span different areas, storage encryption tends to be key. IDC research shows that 77.5% of enterprises have regulatory and/or compliance guidelines that require them to use encryption, and 78.3% of them need to be able to choose from a variety of different encryption approaches to meet their business requirements. When running encryption in line, enterprises are concerned about the performance impact on server and/or storage cycles. Although 60.5% of enterprises predominantly use 128-bit encryption, 52.3% of enterprises have a FIPS 140-2 compliance requirement for 256-bit encryption for at least some workloads.

Many external storage arrays have supported in-line data services such as encryption, compression, and deduplication for years and many use hardware-assist technology to ensure the services do not impact performance. With software-defined solutions like HCI, there is less opportunity for the kind of hardware assist used in arrays. However, Intel Xeon Scalable Processors offer a feature called Intel Quick Assist Technology (QAT) that can provide onboard hardware assist and help address performance concerns with the use of in-line data services such as encryption, compression, and/or deduplication on software-defined infrastructure. Not all server-based storage systems use Intel QAT (the storage software has to be built to specifically take advantage of it), but some do.

Performance density remains a concern, but easy access to NVMe technology should help address the issue. Compared with the SCSI protocol, NVMe offers three orders of magnitude greater parallelism, significantly higher bandwidth, and substantially lower latency. Some NVMe technologies can enable two orders of magnitude lower latency than SCSI. Many 2U24 server-based storage nodes used for HCI configurations, regardless of whether they use Intel or AMD processor technology, can deliver at least 1 million I/O operations per second (IOPS). HCI using NVMe-based nodes excels at delivering low latency and high random read/write IOPS for smaller files, and, in recent years, has become an increasingly popular platform for smaller databases as well.

External storage arrays have tended to lead the industry in the maturity and comprehensiveness of their storage management capabilities. Although early HCI software lacked many of the storage

management capabilities of SAN and NAS systems, that gap has narrowed significantly in recent years. HCI platforms that target general-purpose mixed enterprise workload consolidation, such as Nutanix and VMware, offer almost all the storage management features of external storage arrays, along with agility features not available in those arrays. Stretch cluster configurations that offer a recovery point objective (RPO) of zero tend to still be the province of external storage arrays, but many HCI implementations support snapshot and replication features that enable them to meet stringent RPO and recovery time objective (RTO) requirements.

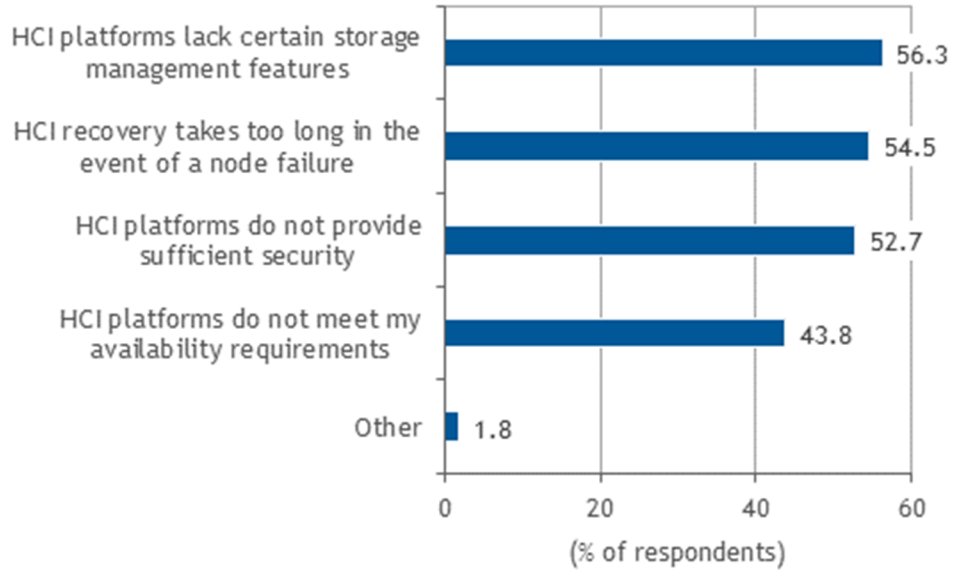
High availability, supported by rapid recovery times, has been a hallmark of external storage arrays for decades. External storage array vendors routinely quote "six-nines plus" availability, and Hitachi, HPE, Infinidat, and NetApp offer a 100% data availability guarantee. Recovery times from device and node failures in HCI configurations has been a stumbling block in the past that has kept HCI vendors from claiming "five-nines" availability. However, HCI vendors have improved their recovery times through use of faster NVMe technology, more widely distributing data through erasure coding that spans cluster nodes and rebuilding only lost data rather than full drives and/or nodes. The fact that 56.6% of enterprises run mission-critical workloads on HCI is an indication that many enterprises think that HCI vendors have come a long way to address the recovery time issue.

To deliver high-availability configurations, most HCI vendors require at least three nodes, and, in smaller configurations, the performance impact of a node loss can be noticeable. Application services can experience significant performance degradation in the event of a node loss, depending on the number of nodes and the heaviness of the workload. The system can enable workloads on failed nodes to automatically restart on other nodes, but there is no getting around the fact that in the wake of a node failure there are fewer compute and storage resources available to service potentially the same workload. External storage arrays tend to implement one of two approaches to handle controller failures: an active/active controller architecture, where both controllers are in use during normal operation and a failure reduces controller resources by half, or an active/passive architecture, where the passive controller acts as a hot standby so there is no performance impact in the event of a controller failure. Active/active architectures tend to be more powerful during normal operation with both controllers contributing to performance, but the impact is potentially large in the event of a failure. Active/passive architectures have customers pay for a controller resource that they rarely use, but they offer a better outcome in the event of a controller failure. One vendor, Infinidat, uses an architecture with three controllers operating in active/active/active mode (see Figure 3).

FIGURE 3

Reasons Not to Put Mission-Critical Workloads on HCI

Q. If you aren't running mission-critical workloads on HCI platforms today, why not?



n = 112

Source: IDC's *HCI Survey,* January 2022

Because HCI can run on virtual machines, it is possible to run a single-node configuration. With the proliferation of edge requirements, it can be helpful to have a single-node option available for customers that need to prioritize cost over high availability for edge workloads. Some vendors offer a two-node HCI configuration option (which also represents an attempt to lower the entry price point for edge and other smaller deployment needs) that offers better availability than single-node implementations. Edge is a rapidly growing market that will drive $25.0 billion in revenue in 2022 and grow to a $41.7 billion market by 2025 (but only a small portion of that will be HCI revenue).

It's interesting to note that within the past year or so, several of the large HCI providers have begun to position themselves as data management vendors or hybrid multicloud providers rather than vendors of HCI platforms. Vendors such as Cohesity, Nutanix, Quantum, Qumulo, and VMware are focusing more on the benefits and capabilities of their HCI platforms rather than positioning themselves as HCI vendors.

The Impact of Digital Transformation and the Move to Cloud Computing

Digital transformation, which is actively under way at 84% of enterprises today, is driving new macrotrends in IT infrastructure. We have moved from an era where most of an enterprise's IT infrastructure resided in its own datacenters to a world where hybrid multicloud is quickly becoming the blueprint. IDC research shows that 64% of enterprises have both an on-premises IT and a cloud component to their IT infrastructure, where the cloud component includes at least two public cloud providers, one of which they may designate as primary. Enterprises are also building private clouds on

premises, and, in the long run, most organizations will choose from three deployment models (traditional IT, private cloud, and public cloud) for workload placement.

As enterprises move through digital transformation, they tend to go through four phases: application and infrastructure modernization, retooling for better automation and orchestration, incorporation of cloud technologies, and migration to a unified management strategy that spans on- and off-premises deployment models. Research shows that 70.0% of respondents believe it is important to have a common underlying architecture between on- and off-premises deployment models, primarily to simplify operations. While a good percentage of public cloud infrastructure is HCI based, an even higher percentage of it is software defined. Software-defined architectures are much more agile than hardware-defined architectures, and agility is a critical requirement for enterprises evolving toward a digital-first orientation. Infrastructure modernization often includes deploying software-defined private cloud infrastructure on premises while also leveraging public cloud infrastructure. IDC's January 2022 *HCI Survey* indicated that 35.3% of survey respondents would use HCI in their private cloud environments.

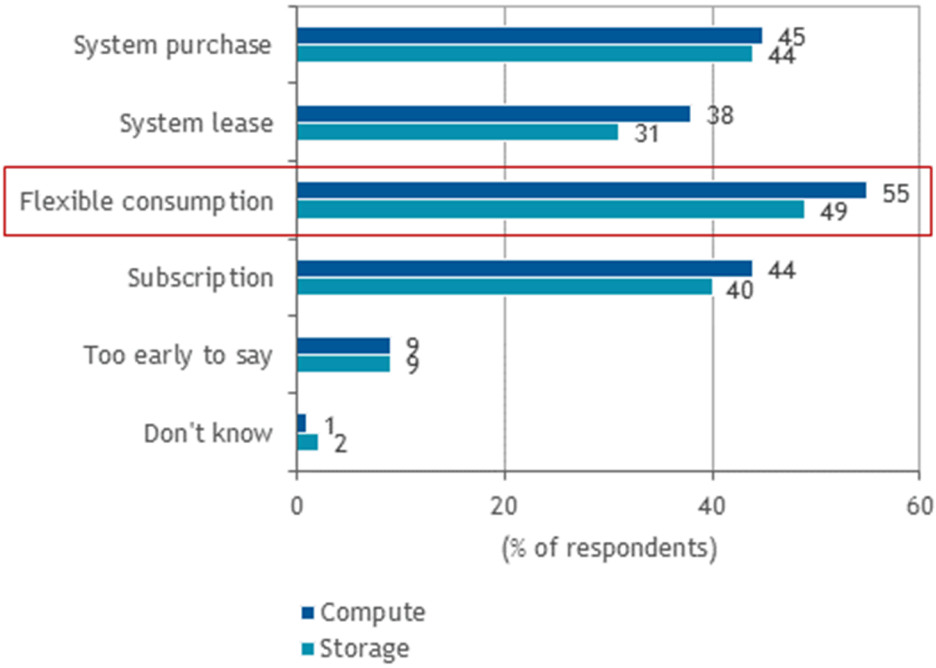
On-premises HCI can provide an easier on-ramp to the public cloud. HCI software that runs on top of commodity hardware can be deployed both on premises and in public clouds, making it easier to migrate virtual machines between the two environments as necessary to service disaster recovery, cloudbursting, and workload balancing, as well as choose the optimal deployment location as applications evolve over their life cycle. Most newly developed applications use a cloud-native, container-based design, and containerizing legacy workloads can help make them more portable as well. HCI is a popular architecture for container-based workloads, with 44.6% of enterprises running containerized workloads on HCI today, and an additional 23% planning to adopt them over the course of 2022. Simple, reliable workload migration capabilities are an important feature in modernized infrastructure.

Flexible consumption is becoming an increasingly important option for enterprises today. While the operational expenditure–based model to finance IT infrastructure initially caught fire with the public cloud, most major enterprise storage vendors now offer a variety of service offerings for on-premises infrastructure deployment, including pay-as-you-go (shown in Figure 4 as flexible consumption) subscription and managed services (not shown in Figure 4). Programs include Dell Technologies APEX, HPE GreenLake, Lenovo TruScale, NetApp Keystone, and Pure Storage Pure as-a-Service. Vendors tend to extend the flexible consumption to most if not all of their infrastructure offerings, giving customers the option to replicate the public cloud experience with their HCI-based on-premises private cloud. Figure 4 indicates that flexible consumption is the most popular deployment model for storage infrastructure going forward, with 49% of respondents selecting it.

FIGURE 4

On-Premises IT Infrastructure Delivery Model Deployment Plans

Q. Which of the following IT infrastructure delivery models does your organization plan to use to deploy compute, storage, and/or data protection resources on premises in 2022?



n = 778

Base = respondents that bought/will buy HCI

Source: IDC's 1Q20 *IT Infrastructure for Compute Survey,* April 2020

There are exciting developments that will unfold in HCI over the next 12 months. In terms of market share by revenue, VMware (with its vSAN-based VMware Cloud Foundation) and Nutanix (with its Nutanix Cloud Platform) lead the market. Projects under development with VMware, Nutanix, and other vendors will:

- Better leverage function accelerator cards to provide hardware offload for a variety of tasks
- Provide better visibility and a more comprehensive set of consistent capabilities from a cloud-based unified management platform
- Enable a more flexible and scalable virtual memory architecture to address the requirements of today's memory-bound workloads (interesting data point: 46.9% of enterprises expect to run into memory-bound limitations with software-defined storage solutions such as HCI)
- Better enable disaggregation of compute and storage resources to expand the types of workloads that can effectively run on top of HCI platforms
- Make it easier to execute dense, mixed workload consolidation with evolving support for block, file, and object storage running on the same HCI platform and managed from a single GUI

Finally, HCI customers are interested in performance-oriented technologies to enable digital transformation, including accelerated compute, in-memory databases, and storage-class memory. IDC research showed that 54.7% of respondents are interested in using accelerated compute (i.e., high-performance GPUs and DPUs available from vendors such as AMD, Fungible, Intel, and NVIDIA), 34.9% want better support for in-memory workloads, and 45.3% want the option to use storage class memory devices (i.e., Intel Optane SSDs, Kioxia FL6) in their HCI nodes.

ADVICE FOR THE TECHNOLOGY BUYER

- Enterprises that think HCI is limited by an inability to scale compute and storage resources independently, insufficient device and/or node recovery times, and inadequate node-level performance and/or capacity density should update their understanding of the capabilities of the rapidly growing, software-defined platform.
- While there are still workloads that are not a good fit for HCI, there are far fewer of them today than there were just two years ago.
- Composable, disaggregated infrastructure (CDI) is in its infancy, but this software-defined approach offers much promise for enterprise infrastructure as it matures. While vendors such as Fungible, GigaIO, Liqid, and TidalScale play in this space today, expect to see Nutanix, VMware, and other HCI vendors make investments and introduce capabilities to better enable CDI on top of their platforms.

LEARN MORE

Related Research

- *Managed VMware Cloud Flex Storage Service Enables Independent Scaling of Storage and Compute Resources on AWS* (IDC #lcUS48996522, March 2022)
- *What to Look for When Considering Enterprise Storage Workload Consolidation* (IDC #US48670822, January 2022)
- *Hyperconverged Infrastructure Adoption Trends – 3Q21: Building Block for Hybrid Cloud Infrastructure* (IDC #US48308121, October 2021)
- *Worldwide Software-Defined Storage Controller Software Forecast, 2021-2025: On-Premises Deployments to Drive Steady Growth in Support of Modern Infrastructure* (IDC #US47404821, October 2021)
- *Worldwide Software-Defined Storage Controller Software Market Shares, 2020: Most Vendors Boost Revenue, APJ Continues Its Sales Growth* (IDC #US47404921, August 2021)

Synopsis

This IDC Perspective provides an overview of recent developments in the hyperconverged infrastructure (HCI) market and encourages enterprises to evaluate these types of platforms as their legacy infrastructure comes up for technology refresh. HCI offerings have come a long way in the past two years and host mission-critical workloads at more than half of all enterprises. Their software-defined agility makes them a popular choice for the infrastructure modernization and digital transformation projects that most enterprises are pursuing.

"Enterprises that still think HCI is limited by an inability to scale compute and storage resources independently, recovery times for device and/or node failures, and node-level performance and

capacity limitations should look again," said Eric Burgener, research vice president, Infrastructure Systems, Platforms and Technologies Group, IDC. "Vendors in this space have come up with new features and approaches that address these concerns, and the best HCI platforms offer a storage management feature set that very closely approximates external storage arrays while differentiating themselves with their flexibility, ease of use, and better economics."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com